Energy functions and their relationship to protein conformation

CS/CME/Biophys/BMI 279, Lecture 3

Sept. 29 and Oct. 1, 2015

Ron Dror

Outline

- Energy functions for proteins (or molecular systems more generally)
  - Definition and properties
  - Molecular mechanics force fields
- What does the energy function tell us about protein conformation?
  - The Boltzmann distribution
  - Microstates and macrostates
  - Free energy

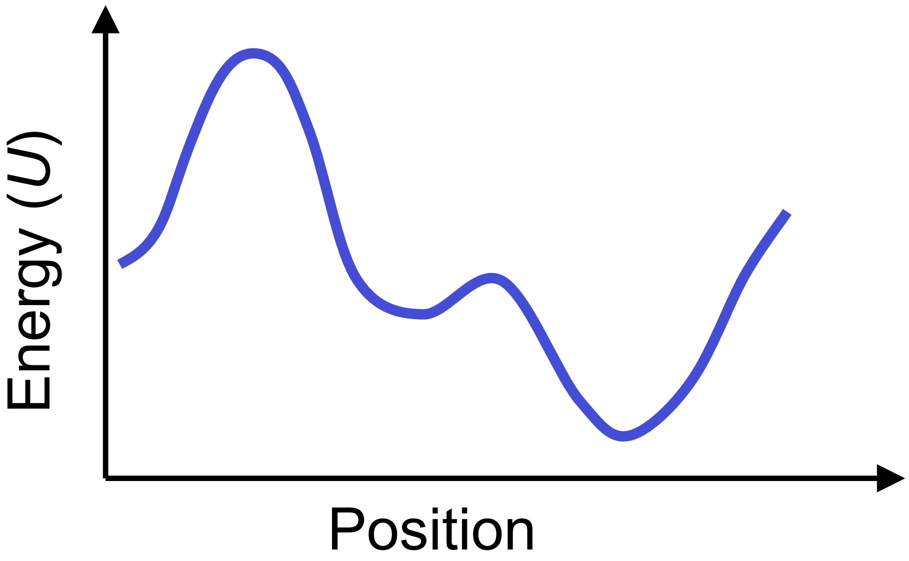
Energy functions for proteins
(or molecular systems more generally)

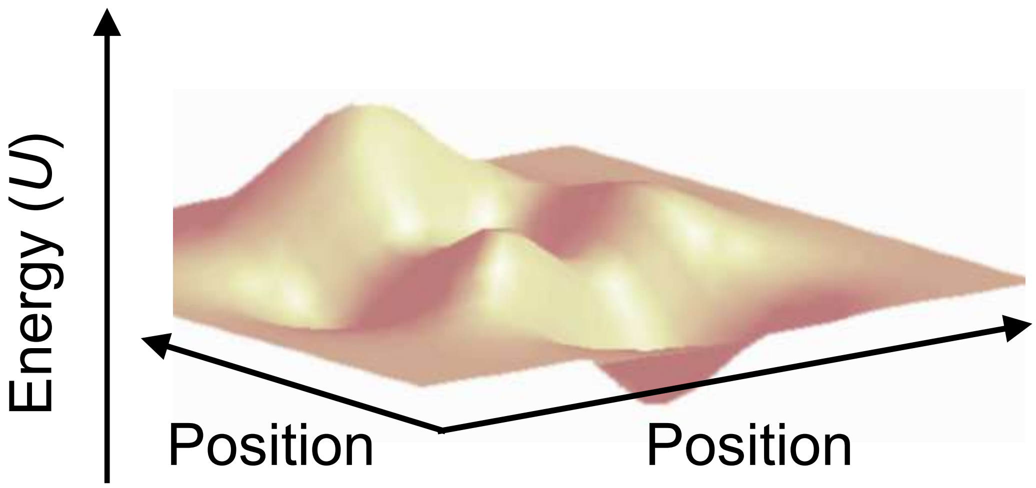
Energy functions for proteins (or molecular systems more generally)

Definition and properties

Energy function

- A potential energy function $U(\boldsymbol{x})$ specifies the total potential energy of a system of atoms as a function of all their positions ($\boldsymbol{x}$)
  - For a system with n atoms, $\boldsymbol{x}$ is a vector of length $3n$ (x, y, and z coordinates for every atom)
  - In the general case, include not only atoms in the protein but also surrounding atoms (e.g., water)



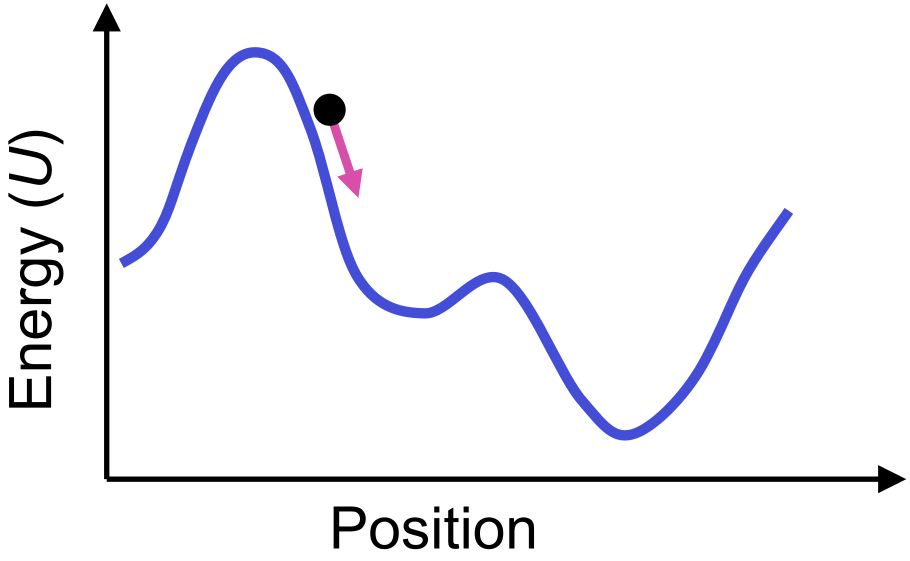


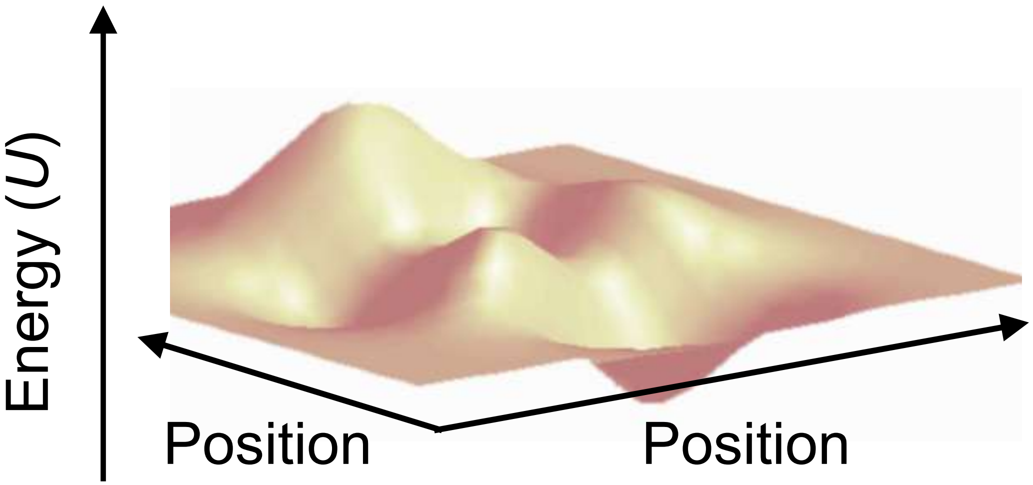
Relationship between energy and force

- Force on atom *i* is given by derivatives of *U* with respect to the atom's coordinates x_i, y_i, and z_i

$$F(\boldsymbol{x}) = -\nabla U(\boldsymbol{x})$$

- At local minima of the energy *U*, all forces are zero
- The potential energy function *U* is also called a *force field*





Types of force fields (energy functions)

- A wide variety of force fields are used in atomic-level modeling of macromolecules
- Physics-based vs. knowledge-based
  - Physics-based force fields attempt to model actual physical forces
  - Knowledge-based force fields are based on statistics about, for example, known protein structures
  - Most real force fields are somewhere in between
- Atoms represented
  - Most realistic choice is to model all atoms
  - Some force fields omit waters and other surrounding molecules. Some omit certain atoms within the protein.

Energy functions for proteins (or molecular systems more generally)

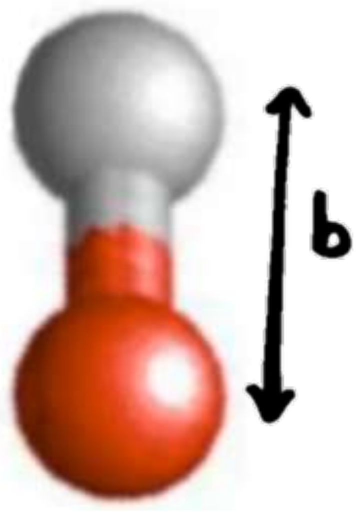
Molecular mechanics force fields

Molecular mechanics force fields

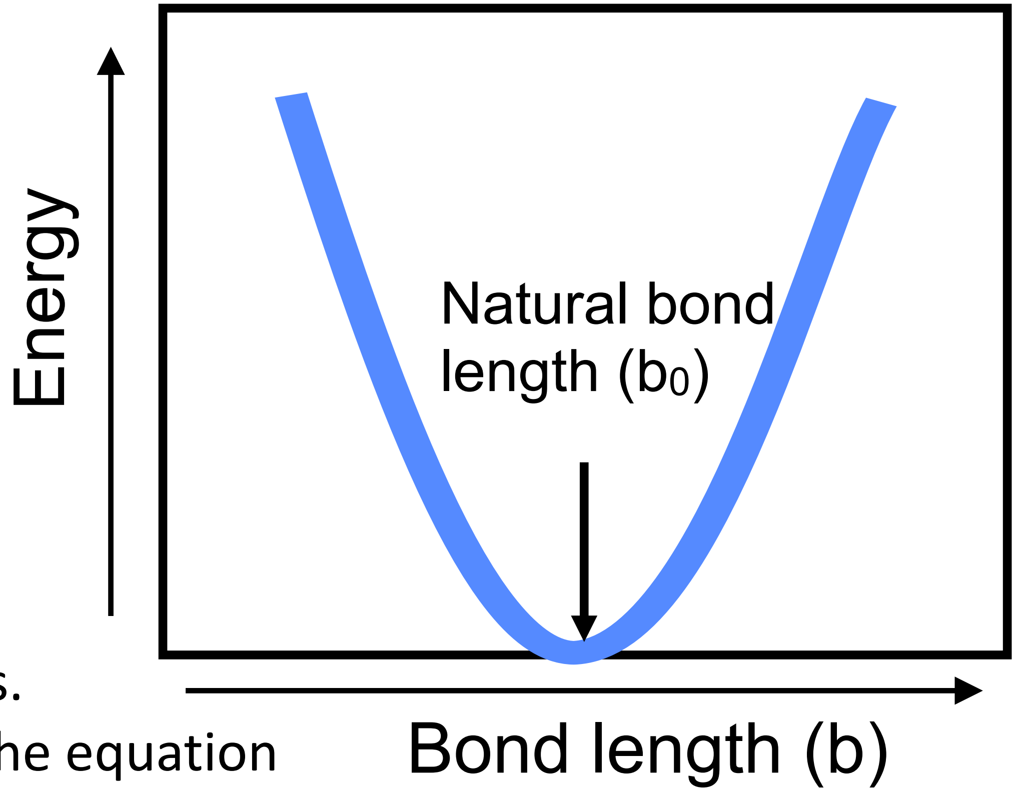
- Today, we'll focus on *molecular mechanics force fields*, which are often used for molecular simulations
- These are more toward the physics-based, all-atom end (i.e., the more "realistic" force fields)
  - Represent physical forces explicitly
  - Typically represent solvent molecules (e.g., water) explicitly
- We'll revisit the forces acting between atoms and write down the functional forms typically used to approximate them

Bond length stretching

- A bonded pair of atoms is effectively connected by a spring with some preferred (natural) length. Stretching or compressing it requires energy.



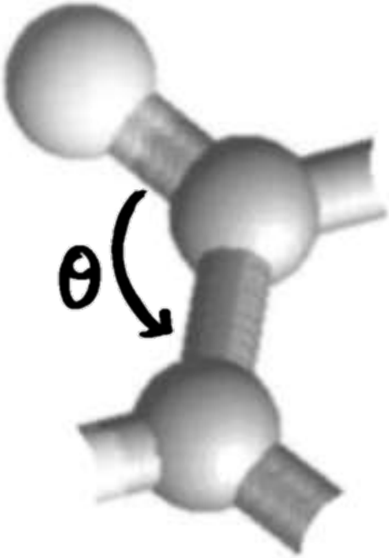
$$U(b) = k_b\left(b - b_0\right)^2$$



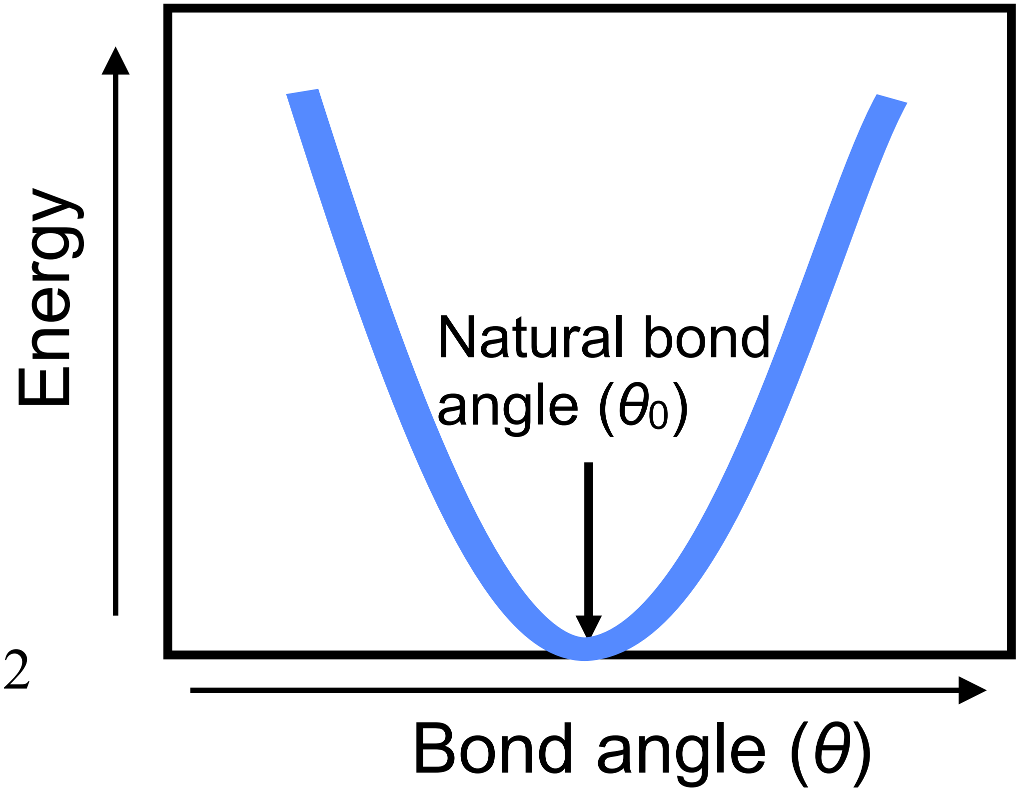
Note: I'm folding constants into units.
Often, a factor of 1/2 is included in the equation

Bond angle bending

- Likewise, each bond angle has some natural value. Increasing or decreasing it requires energy.

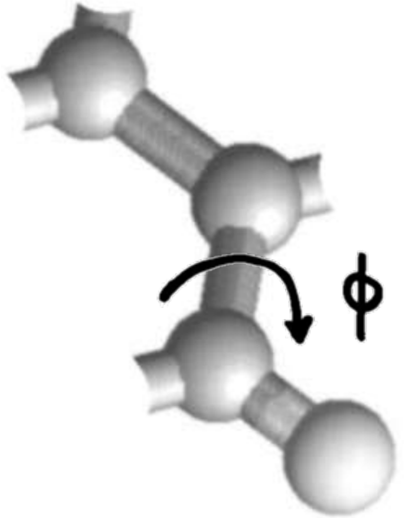


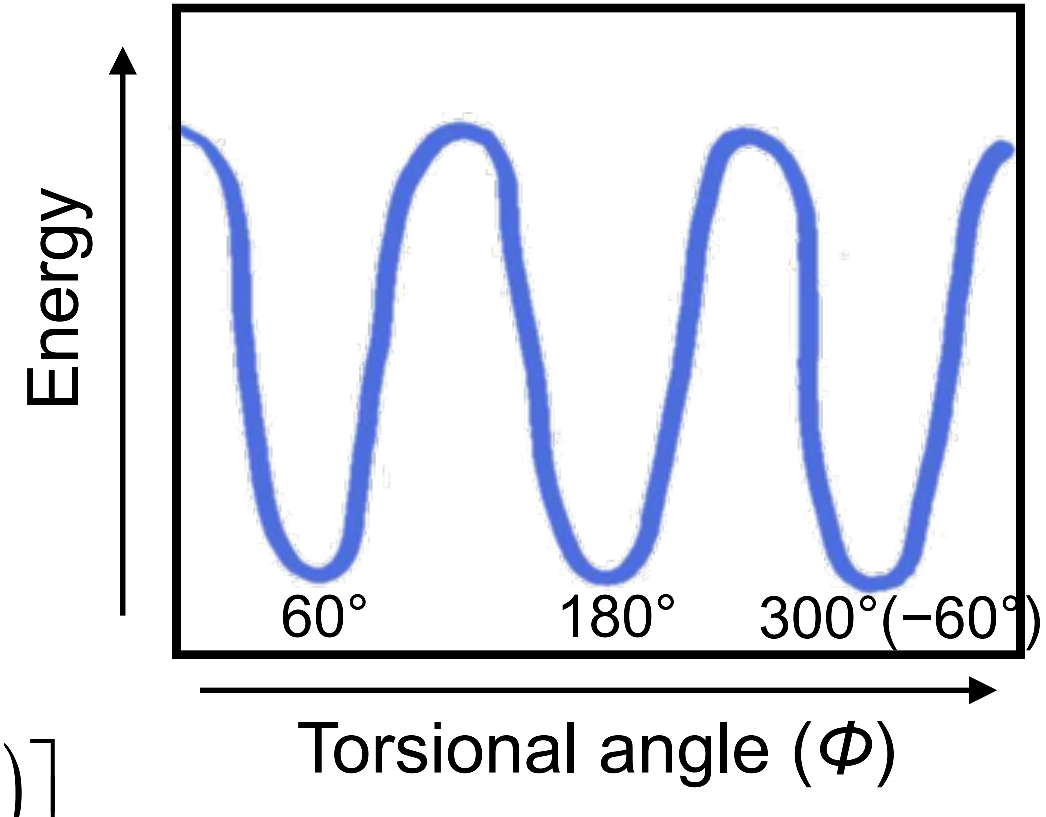
$$U(\theta) = k_\theta \left(\theta - \theta_0\right)^2$$



Torsional angle twisting

- Certain values of each torsional angle are preferred over others.

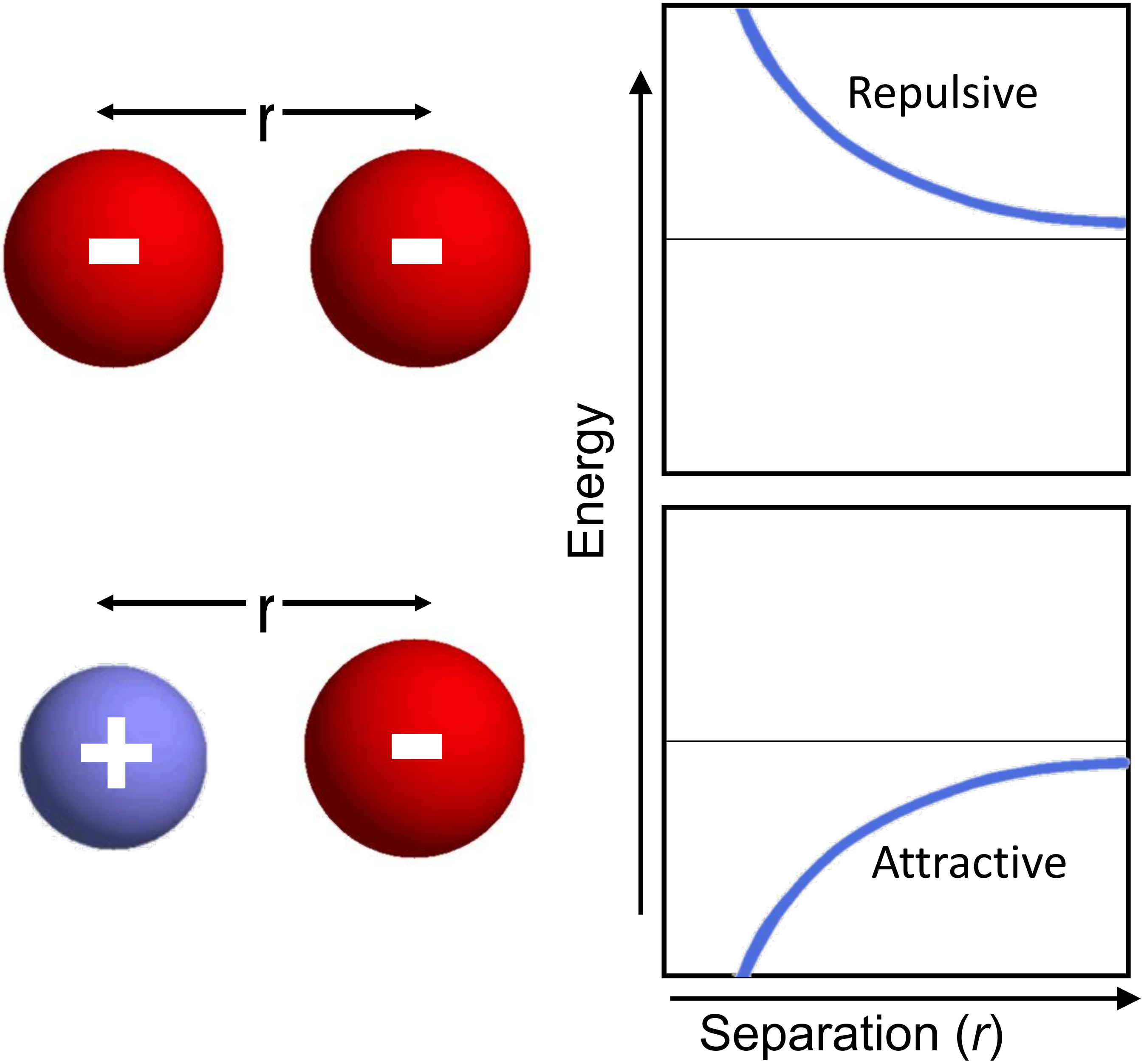




$$U(\phi) = \sum_n k_{\phi,n}\left[1+\cos\left(n\phi-\phi_n\right)\right]$$

Typically n takes on one or a few values between 1 and 6 (particularly 1, 2, 3, 6)

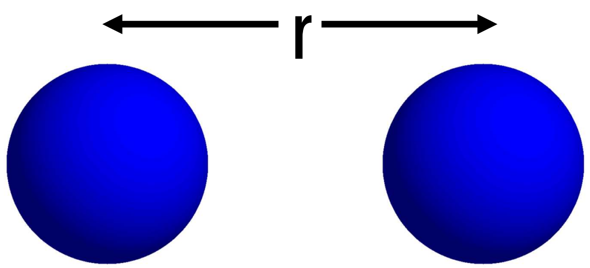
Electrostatics interaction



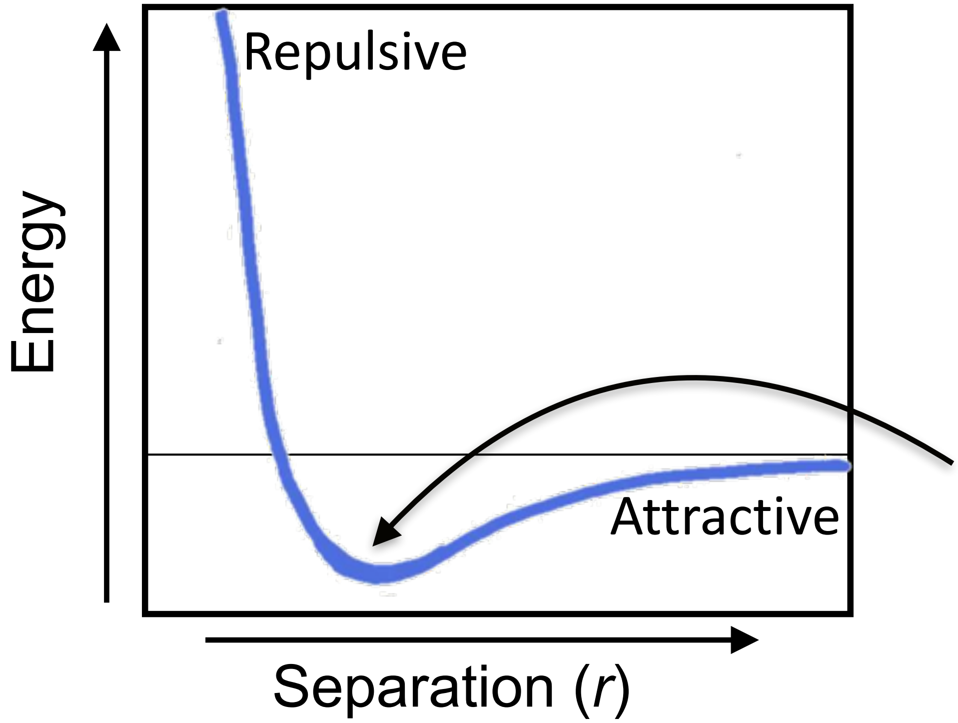
- Like charges repel. Opposite charges attract.
- Acts between all pairs of atoms, including those in different molecules.
- Each atom carries some “partial charge” (may be a fraction of an elementary charge), which depends on which atoms it’s connected to

$$U(r) = \frac{q_i q_j}{r}$$

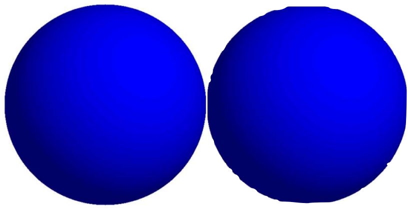
van der Waals interaction



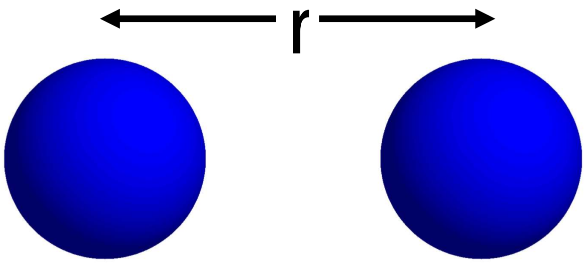
- van der Waals forces act between all pairs of atoms and do not depend on charge.
- When two atoms are too close together, they repel strongly.
- When two atoms are a bit further apart, they attract one another weakly.



Energy is minimal when atoms are "just touching" one another



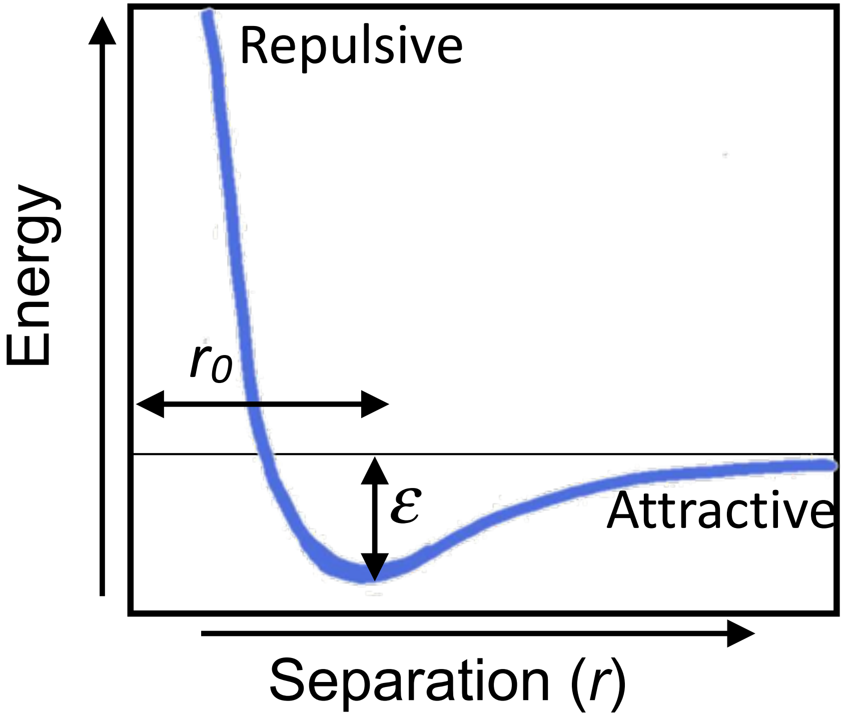
van der Waals interaction



$$U(r) = \frac{A_{ij}}{r^{12}} - \frac{B_{ij}}{r^{6}}$$

We can also write this as:

$$U(r) = \varepsilon\left[\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^{6}\right]$$



Note: Historically, r^{12} term was chosen for computational convenience; other forms are sometimes used

A typical molecular mechanics force field

$$U = \sum_{\text{bonds}} k_b \left(b - b_0\right)^2$$

Bond lengths (“Stretch”)

$$+ \sum_{\text{angles}} k_\theta \left(\theta - \theta_0\right)^2$$

Bond angles (“Bend”)

$$+ \sum_{\text{torsions}} \sum_{n} k_{\phi,n} \left[1 + \cos\left(n\phi - \phi_n\right)\right]$$

Torsional/dihedral angles

Bonded terms

$$+ \sum_{i} \sum_{j>i} \frac{q_i q_j}{r_{ij}}$$

Electrostatic

$$+ \sum_{i} \sum_{j>i} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}$$

Van der Waals

Non-bonded terms

How are the parameters fit?

- Combination of:
  - Quantum mechanical calculations
  - Experimental data
    - For example: b_0 can be estimated from x-ray crystallography, and K_b from spectroscopy (infrared absorption)

$$U(b) = K_b\left(b - b_0\right)^2$$

- The torsional parameters are usually fit last. They absorb the “slop.” Fidelity to physics is debatable.
- These force fields are approximations!

What does the energy function tell us about protein conformation?

What does the energy function tell us about protein conformation?

The Boltzmann distribution

Relating energy to probability

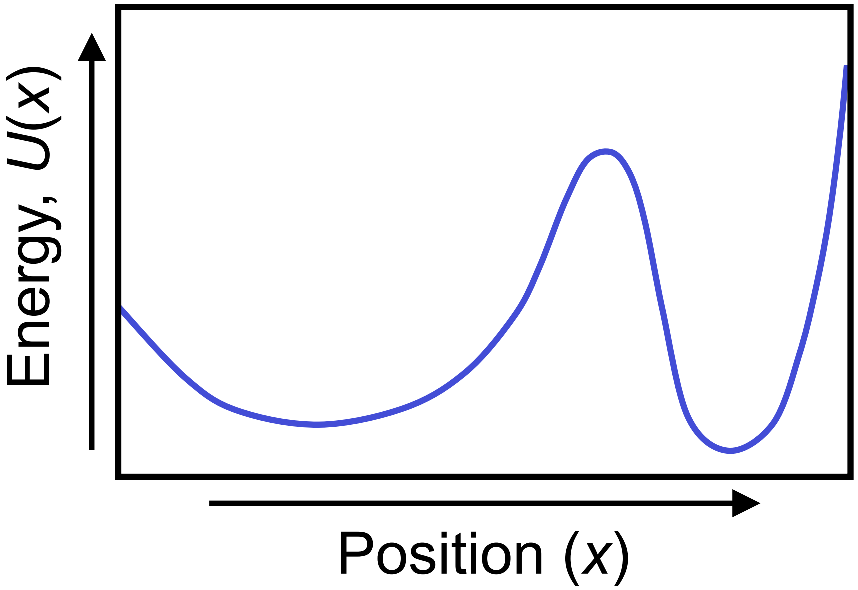
- Given the potential energy associated with a particular arrangement of atoms (set of atom positions), what is the probability that we'll see that arrangement of atoms?
- Assumptions:
  - System is at constant temperature. Atoms are constantly jiggling around.
  - We watch the system for a really long time (allow it to fully equilibrate).

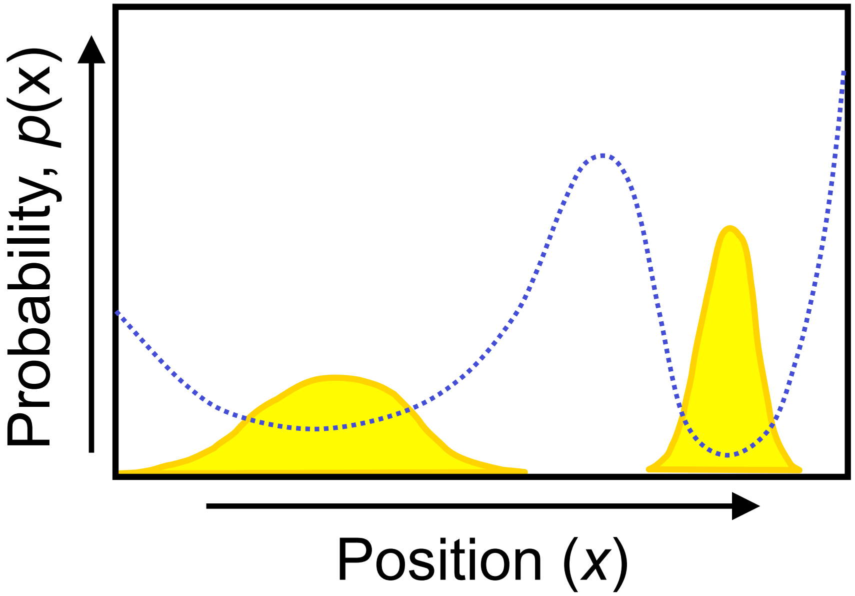
The Boltzmann Distribution

- The Boltzmann distribution relates potential energy to probability

$$p(\boldsymbol{x}) \propto \exp\left({}^{-U(\boldsymbol{x})}\!\big/_{k_B T} \right)$$

where T is temperature and k_B is the Boltzmann constant

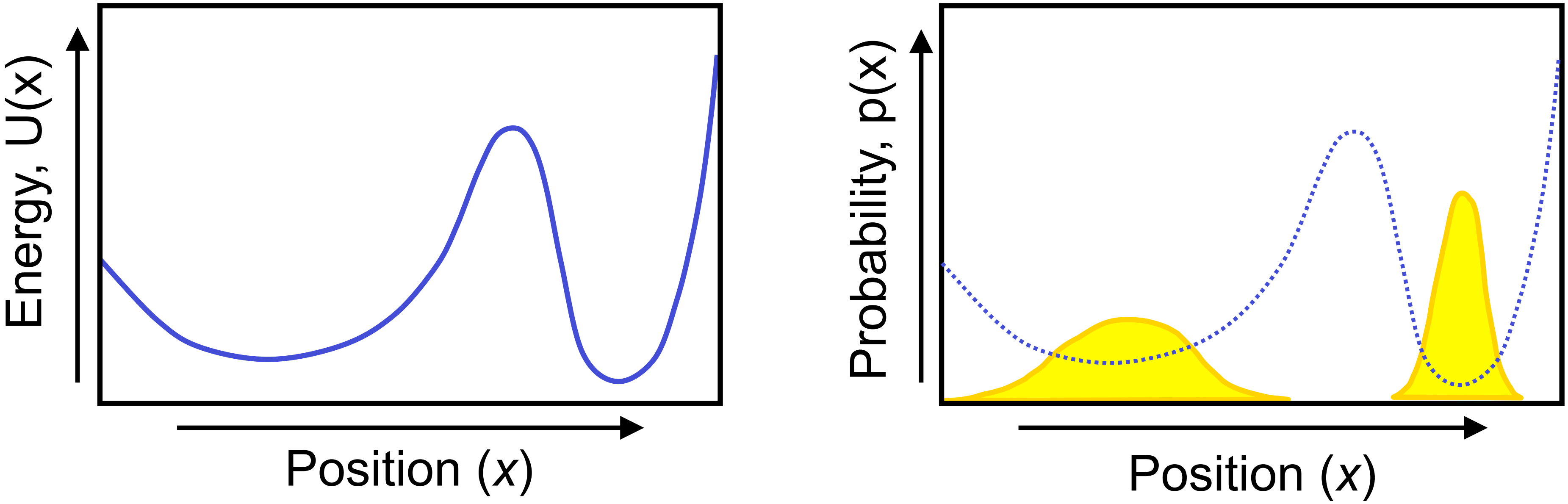




The Boltzmann Distribution

- Key properties:
  - Higher energy gives lower probability
  - Exponential relationship: each time probability halves, energy increases by a constant
  - Temperature dependence: at higher temperature, need to increase energy more for same probability reduction

$$p(\boldsymbol{x}) \propto \exp\left({-U(\boldsymbol{x})} \big/ {k_B T} \right)$$



What does the energy function tell us about protein conformation?

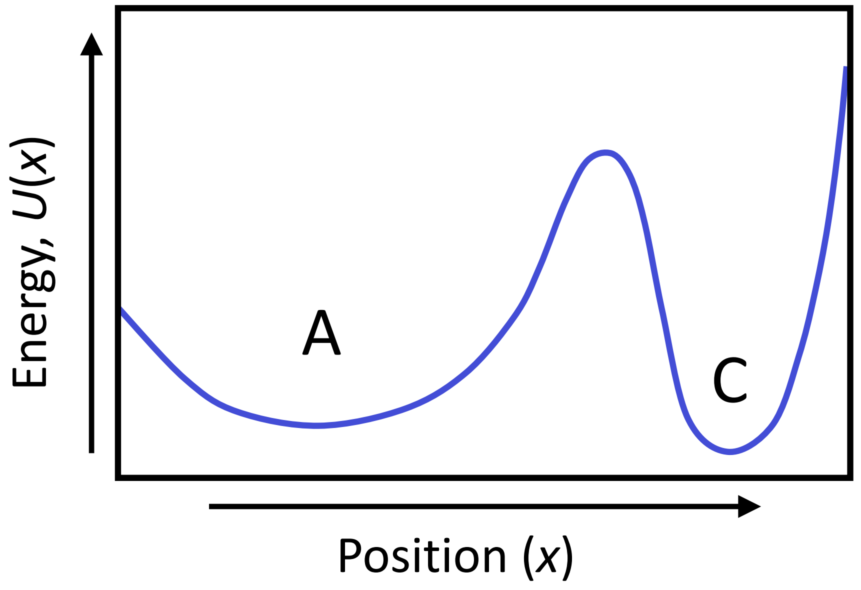
Microstates and macrostates

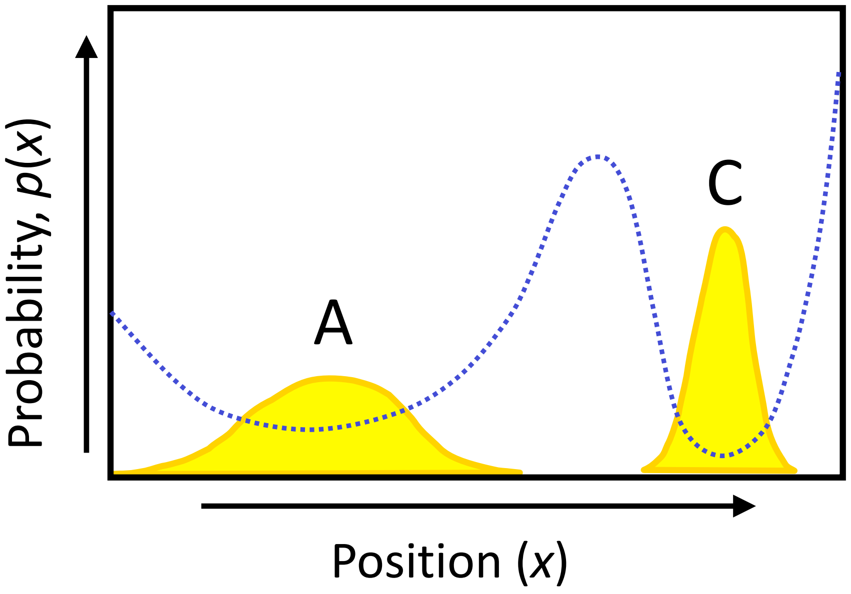
Protein structure: what we care about

- We don’t really care about the probability that all the atoms of the protein and all the surrounding water atoms will be in one precise configuration
- Instead, we care about the probability that protein atoms will be in some *approximate* arrangement, with *any* arrangement of surrounding water

Protein structure: what we care about

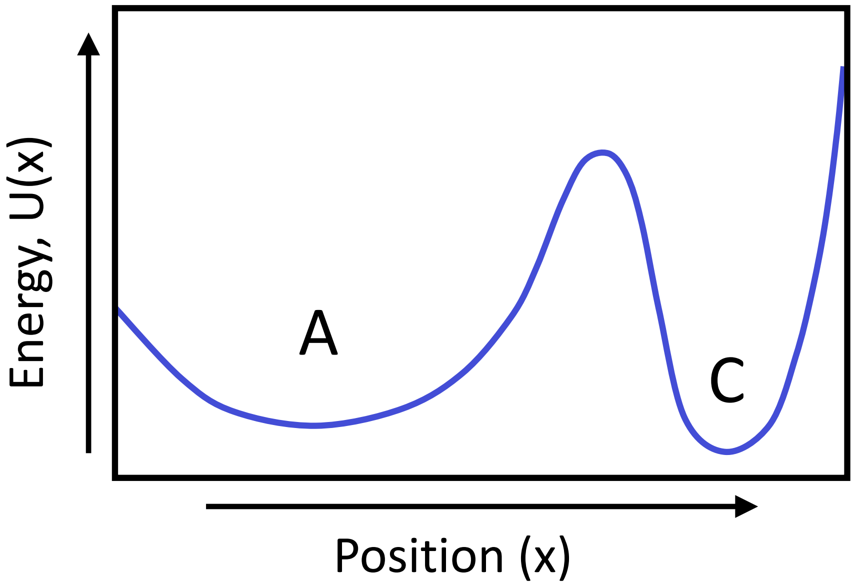
- In other words, we wish to compare different sets (neighborhoods) of atomic arrangements
- We define each of these sets as a *macrostate* (A, C). Each macrostate includes many *microstates*, or specific atom arrangements ***x***.
  - Macrostates—also called conformational states—correspond to wells in energy landscape

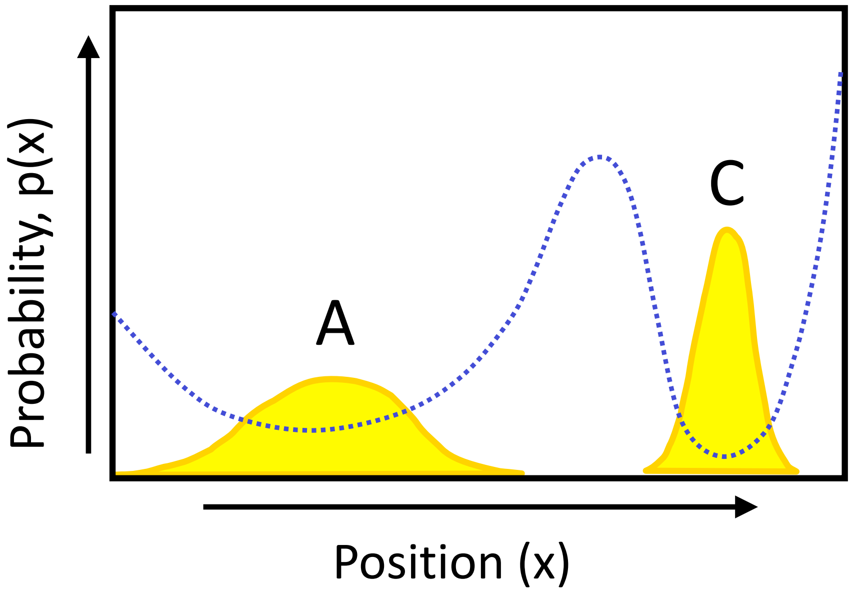




Probabilities of macro states

- Which has greater probability, A or C?
  - C is a deeper well, so individual atomic arrangements more likely
  - A is a broader well, so it includes more distinct individual arrangements



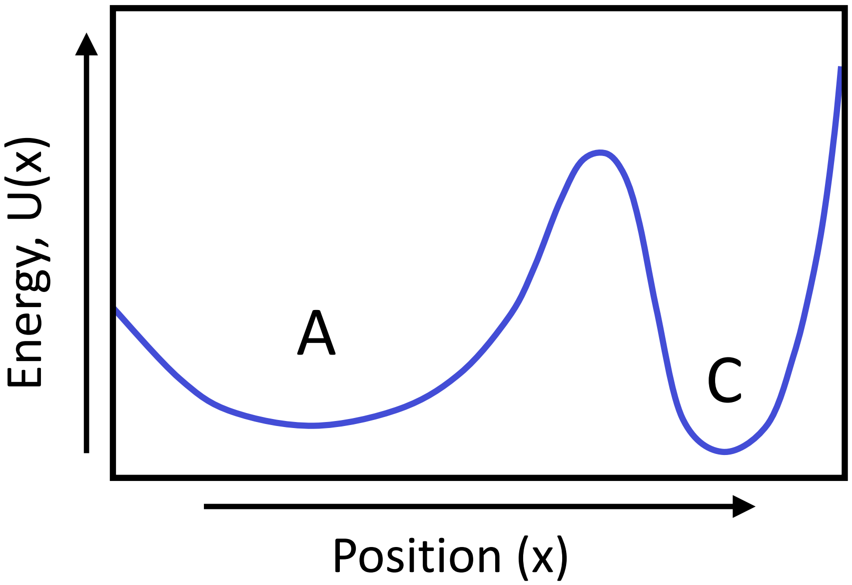


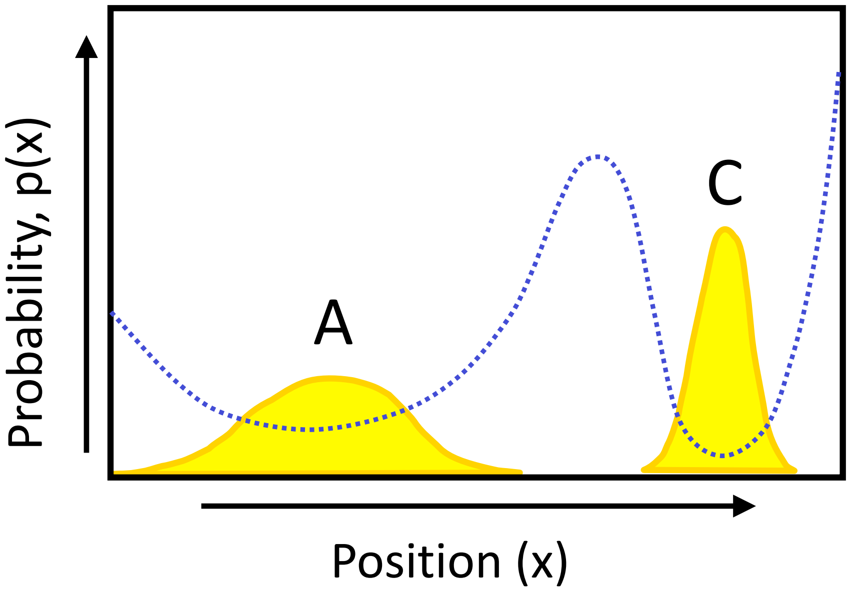
Probabilities of macro states

- Which has greater probability, A or C?
- To get probability of macrostate, sum/integrate over all microstates within it

$$P(A) = \int_{\boldsymbol{x}\in A} P(\boldsymbol{x}) \propto \int_{\boldsymbol{x}\in A} \exp\left({-U(\boldsymbol{x})}\big/{k_B T} \right) d\boldsymbol{x}$$

- At low temperature, P(C) > P(A)
- At high temperature, P(A) > P(C)





What does the energy function tell us about protein conformation?

Free energy

Free energy of a macro state

- So far we have assigned energies only to microstates, but it's useful to assign them to macrostates as well.
- Define the *free energy* G_A of a macrostate A such that:

$$P(A) = \exp\left({-G_A}/{k_B T} \right)$$

- This is analogous to Boltzmann distribution formula:

$$p(\boldsymbol{x}) \propto \exp\left({-U(\boldsymbol{x})}/{k_B T} \right)$$

Free energy of a macro state

- Define the *free energy* G_A of a macrostate A such that:

$$P(A) = \exp\left({-G_A} \middle/ {k_B T} \right)$$

- Solving for G_A gives:

$$G_A = -k_B T \log_e \left(P(A) \right)$$

- One can also express free energy in terms of enthalpy (mean potential energy, *H*) and entropy (“disorder”, *S*):

$$G_A = H_A - TS_A$$

You’re not responsible for this last equation, or for the definitions of enthalpy and entropy

So which conformational state will a protein adopt?

- The one with the *minimum free energy*
  - Wide, shallow wells often win out over narrow, deep ones
- This depends on temperature
- At room or body temperature, this is very different from minimum potential energy